E C O N O M I C S   B U L L E T I N

A consistent nonparametric estimation of spatial autocovariances

Théophile Azomahou
BETA, Universit?Louis Pasteur, Strasbourg 1

Dong Li
Department of Economics, Kansas State University

Abstract

We examine some aspects of estimating sample autocovariances for spatial processes. Especially, we note that for such processes, it is not possible to approximate the expectation by the sample mean, like in the case of time series data. Then, we propose a consistent nonparametric estimation of sample autocovariances for an irregularly scattered spatial process, derived from a transformation of the initial process. We also suggest an L_2−consistent weighting matrix. Monte Carlo simulations are used to evaluate the performance of the proposed estimators in finite samples.

Citation: Azomahou, Théophile and Dong Li, (2005) "A consistent nonparametric estimation of spatial autocovariances." *Economics Bulletin,* Vol. 3, No. 29 pp. 1−10
Submitted: April 14, 2005. **Accepted:** June 1, 2005.
URL: http://www.economicsbulletin.com/2005/volume3/EB−05C10005A.pdf

1 Motivation

General forms of dependence are rarely allowed for in cross-sectional data, although routinely permitted in time series. A major reason is that time series typically come with a time label giving the data a natural ordering and a structure which is absent in cross-sectional data. In practice, this makes estimation and inference more difficult for cross-sectional data with dependence than that for time series.

Spatial dependence is defined as a special case of non-zero covariance structure for cross-sectional observations at pairs of locations. Contrary to time series models, until quite recently, little attention has been paid to models for spatial data. The problem follows from the fact that spatial processes are characterized by a system of dependence which, by nature, involves multi-directional motion whereas the dependence in time is uni-directional. This particular characteristic of spatial stochastic processes precludes the simple transposition of time series methodologies as noticed by Anselin and Bera (1998).

Recently several studies have developed specific methods to estimate models based on spatial data. Kelejian and Prucha (1998, 1999) suggested a generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance. Conley (1999) proposed a generalized method of moments estimation based on the concept of agents' economic distances. The study of Sajjan (2000) extended the m-dependence concept to two-dimensional lattice processes, and derived central limit theorems for such processes via generalization of time series methods. Chen and Conley (2001) developed a semiparametric spatial framework using panel data models. The present paper is more in line with the work of Driscoll and Kraay (1998) which stated a consistent covariance matrix estimation with spatially dependent panel data. They presented some mixing conditions under which a simple extension of common nonparametric covariance matrix estimation techniques yields standard error estimates that are robust to very general forms of individual and serial dependence as the time dimension becomes large.

However, our study differs from that of Driscoll and Kraay (1998) in two points. First, the GMM estimator of Driscoll and Kraay (1998) is based on panel data and the model is identified using the cross-sectional averages of the transformation of the orthogonality conditions. As a result, their asymptotic approximation is based on $\sqrt{T}$-consistency. Here, we are interested in estimating spatial autocovariances in a purely cross-section framework. Our results can be extended to panel data. Unlike Driscoll and Kraay (1998), our results are applicable even if T is small. We present conditions under which consistency can be obtained via a simple transformation of the initial spatial process. This implies that spatial covariates can be identified from the n spatial data available without having to specify the form of the spatial dependence. Secondly, we show that, based on this consistent estimation of the sample autocovariances, one can derive easily a consistent estimator of a weighting matrix.

First let us start with a time series example. Estimation procedures developed by, among others, Hansen (1982) and Domowitz and White (1984) allow to compute GMM estimators based for example on time series data. These procedures make use of an orthogonality condition $E[h_t(\theta)] = 0$, where θ is a $k \times 1$ vector of parameters to be estimated, and $h_t(\theta)$ is a $r \times 1$ vector of functions of the data and parameters, where $r \geq k$. The GMM estimator proposed by Hansen (1982) is obtained by choosing θ such that

$$\hat{\theta}_{GMM} = \arg\min_{\theta \in \Theta} h_T(\theta)' \hat{W}_T h_T(\theta), \tag{1}$$

where $h_T(\theta) = 1/T \sum_{t=1}^{T} h_t(\theta)$ defines the vector of sample moments and $\hat{W}_T$ is a random, symmetric weighting matrix. Hansen (1982) and Domowitz and White (1984) have showed that the asymptotic covariance matrix of $\hat{\theta}_{GMM}$ is given by

$$\psi_T = (H_T' W_T H_T)^{-1} H_T' W_T S_T W_T H_T (H_T' W_T H_T)^{-1}, \tag{2}$$

where $S_T = 1/T \sum_{s=1}^{T} \sum_{t=1}^{T} E[h_s(\hat{\theta}) h_t(\hat{\theta})']$, $H_T = 1/T \sum_{t=1}^{T} E[\frac{\partial}{\partial \theta} h_t(\hat{\theta})]$ is the $(r \times k)$ matrix of partial derivative and W_T is a nonrandom matrix such that plim $(\hat{W}_T - W_T) = 0$. We know that a consistent estimation of the asymptotic covariance matrix is important for the computation of asymptotic confidence intervals and hypothesis testing. It is also known that estimation of S_T is difficult and is also more important since the formation of an optimal GMM estimator required that $\hat{W}_T$ is a consistent estimator of S_T^{-1}, as demonstrated by Hansen (1982). The simplest estimator of S_T takes the form

$$\hat{S}_T = \hat{\Omega}_0 + \sum_{j=1}^{m} \left[\hat{\Omega}_j + \hat{\Omega}_j' \right], \quad \hat{\Omega}_j = \sum_{t=j+1}^{T} h_t(\hat{\theta}) h_{t-j}(\hat{\theta}),$$

where the bound m is the number of sample autocovariances $\hat{\Omega}_0$ used to form $\hat{S}_T$.

Time domain techniques to compute S_T that is positive semi-definite suggest the use of spectral procedures and are motivated by the fact in the covariance stationary case, the limit of S_T is 2π times the spectral density of $h_t(\hat{\theta})$ at frequency zero. Although these procedures are difficult to apply in practice, they are technically possible in time series framework, but not in spatial framework. So, our focus in this paper is to obtain a spatial analogue of $\hat{\Omega}_j$ and $\hat{S}_T$. These estimators are easy to compute in practice and robust to misspecifications of the underlying spatial processes.[1]

In the remaining of this paper, In Section 2 we propose a nonparametric estimation of an irregularly scattered spatial process. We show that the suggested estimator is consistent in probability. Then we derived a consistent weighting matrix, which is proved to converge in norm L_2. In Section 3 the Monte Carlo simulations are carried out to evaluate the performance of the proposed estimator.

2 Estimation of Spatial Autocovariance

Contrary to time series, there is typically no natural order for arranging a spatial sample, due to its multi-directional motion. The spatial autocovariance function cannot then be approximated from the sample analogue as is the case for time series. We will

[1]It should be noted that to form a spatial GMM estimator, orthogonality conditions for spatial stochastic processes are required. Examples of such conditions can be found, among others, in Kelejian and Prucha (1998, 1999) and Conley (1999).

outline the reason later. At this stage, we have two possibilities. First, we can suppose that the underlying data generating process follows from regularly scattered spatial sites. However, in practice spatial data rarely exhibit such a characteristic. Instead, we consider the case of irregularly scattered spatial data which seems more natural.

Let $s \in \mathbb{R}^d$ be a generic data location in a d-dimensional Euclidean space and $\{U(s), s \in \mathbb{D} \subset \mathbb{R}^d\}$ be a second-order stationary spatial process with zero mean and finite variance.

Definition 1 *For a distance lag $\tau \in \mathbb{R}^d$, we define the autocovariance function as*

$$C_n(\tau) = cov\left[U(s+\tau), U(s)\right] \equiv E\left[U(s+\tau)U(s)\right]. \tag{3}$$

Later in Assumption 1 we will explicitly assume that $U(.)$ is a stationary process in the sense that its autocovariance is a function of the distance but not the locations. This is similar to the covariance stationarity in time series. See Cressie (1991) for more details.

For a given distance lag τ, one wishes to obtain a consistent estimation of the spatial autocovariance function $C_n(\tau)$ associated with $U(s)$. In the sequel, index n will be omitted to ease presentation when it is not necessary.

Definition 2 *From $U(s)$, let us define the random variable $Z(\tau)$ as*

$$Z(\tau) = U(s+\tau)U(s). \tag{4}$$

The cloud of points denoted N_U constructed from the n realizations $u(s)$ of $U(s)$ allows us to define a cloud of points denoted N_Z associated with the $n(n-1)/2$ realizations $z(\tau)$ of $Z(\tau)$ by setting

$$z(\tau) = u(s+\tau)u(s). \tag{5}$$

The difficulty here is that one has only one observation for a distance τ between two points of the sample. It is then not possible to construct an estimator of $E[Z(\tau)]$ as an empirical mean of a large sample like in the case for time series data. To overcome this issue, we propose to fit the cloud of points N_Z by a smooth function which will be a consistent approximation of relation (3). The intuition is the following.

Assuming $C_n(\tau)$ to be regular and the sample points to be uniformly distributed, $C_n(\tau)$ can be estimated by smoothing the scatter plot N_Z using the kernel method. Intuitively, the smoothing function should provide a consistent nonparametric estimation for $C_n(\tau)$ which is then robust to spatial correlation of unknown form. Several assumptions are needed to construct this estimator and to prove its consistency.

Assumption 1 *$\{U(s), s \in [a,b]\}$ is a second-order stationary spatial process with zero mean. Furthermore, assume that there exists a spatial autocovariance function $C_n(\tau)$ for $U(s)$ such that $C_n(\tau)$ is $\mathcal{C}^\infty$.*

$$E(U(s)) = 0, \text{ for all } s \in \mathbb{D}, \tag{6}$$

and $E(U(s)U(s+\tau))$ is a function of the distance τ but not the location s, and thus can be written as $C_n(\tau)$. Furthermore, we assume that

$$C_n(\tau) \to 0 \text{ as } |\tau| \to \infty. \tag{7}$$

Assumption 2 *We define the following sets:*

(i) $I = \{i_1, \cdots, i_n\}$ *is a set of* n *coordinates with* i_j *having a uniform distribution* $\mathcal{U}_{[a,b]}$*,*

(ii) $\Lambda = \{\tau = |i - i'|/(i, i') \in I\}$ *with* $\max\{0, \tau - \Delta(n)\} < i - i' < \Delta(n) + \tau$*, where* $\Delta(n)$ *denotes the maximum distance or the maximum allowable error,*

(iii) $\Psi_n^\tau = \{(i, i') \in I \times I/\, |i - i'| \in B(\tau, \Delta(n)), \tau \in \Lambda\}$*, where* $B(\tau, \Delta(n))$ *is an open ball, and* n *here indicates the number of coordinates,*

(iv) $\lim_{n\to\infty} \Delta(n) = 0$.

Assumption 3 *There exists a function* K *defined onto* $[0, \Delta(n)]$*, positive, strictly decreasing, with* $K(0) = 1$ *and* $K(\Delta(n)) = 0$.

Proposition 1 *Given Assumptions 1–3, a consistent nonparametric estimator* $\hat{C}_n(\tau)$ *of* $C_n(\tau)$ *is such that*

$$\hat{C}_n(\tau) = \sum_{(i,i')\in\Psi_n^\tau} K\left(||i - i'| - \tau|\right) U(i)U(i') \xrightarrow[n\to\infty]{p} C_n(\tau). \tag{8}$$

Proof. See the appendix.

The estimator (8) is nonparametric in that the form of the spatial correlation between the $U(s)$ at different τ does not need to be specified. The interest of this result is that it provides a consistent estimation of spatial covariance of unknown form directly from the n data points. Note how this contrasts with a parametric framework. Indeed, in such a situation, one might estimate $n(n-1)/2$ parameters from n data points. Since it is impossible to estimate parametrically such covariance terms directly from n data points due to identification issues. It is then necessary to impose sufficient constrains on the spatial interactions such that a finite number of parameters characterizing the correlation can be consistently estimated. This is usually achieved in parametric framework by posing a spatial contiguity matrix; see e.g. Anselin and Bera (1998) for a nice survey.

Once we have $\hat{C}_n(\tau)$ at hand, a consistent estimation of the weighting matrix, say, V_n can be formed easily using for example nearest neighbor procedures ($k - nn$). See e.g. Härdle (1990) for details on the $k - nn$ method.

First of all, note from (8) that $\hat{C}_n(\tau)$ is symmetric by construction. The weights are defined as ω_{ij} for $i \neq j$ with the restriction of uniform boundedness. Then a consistent estimator $\hat{V}_n$ of V_n is

$$\hat{V}_n = \hat{C}_n(0) + \frac{1}{2} \sum_{\substack{(i,j)\in\Psi_n^\tau \\ j\neq i}} \omega\left(||i - j| - \tau|\right) \hat{C}_n(\tau), \tag{9}$$

with $\hat{C}_n(0)$ denoting the variance at zero lags. We may expect that the estimator $\hat{V}_n$ formed by smoothing sample autocovariances with weights $\omega(.)$ that approach one as $n \to \infty$, should be consistent. The consistency of (9) is stated below.

Proposition 2 *Given Assumptions 1–3 and Proposition 1*

$$\left\{ \hat{C}_n(0) + \frac{1}{2} \sum_{\substack{(i,j)\in\Psi_n^\tau \\ j\neq i}} \omega\left(||i-j|-\tau|\right) \hat{C}_n(\tau) \right\} - V_n \xrightarrow[n\to\infty]{L_2} 0. \tag{10}$$

Proof. See the appendix.

3 Monte Carlo Simulations

This section presents the results from a small Monte Carlo (MC) simulation that examines the finite sample properties of lour estimator. The experiment uses the simplest possible generating data process (GDP) where location and distances are fixed. Firstly, data are generated from a gaussian process spatially correlated $Y(s)$. Secondly, from these data, we compute the realization associated to the variable $Z(h) = Y(s)Y(s+h)$. Finally, the cloud of points obtained is smoothed using the kernel method. Note that to simplify, the simulation uses one dimensional spatial process ($s \in \mathbb{R}$).

We assume that the data are uniformly distributed. The first step consists in generating the vector of coordinate x (sorted in increasing order) of points y_i of the sample. Then, we define the matrix D of distances between different points as

$$D = \begin{bmatrix} 0 & |x_1 - x_2| & ... & |x_1 - x_n| \\ |x_2 - x_1| & & & \vdots \\ \vdots & & & \vdots \\ |x_n - x_1| & ... & ... & 0 \end{bmatrix}$$

The matrix of spatial correlation above is computed by assuming that the spatial dependence is of the form

$$cov(y_1, y_2) = \exp\{-\alpha\,|x_1 - x_2|\}$$

which is computed the above by setting set $\alpha = 0.5$. The vector of observations for the variable of interest is then generated as follows. In a first step, a gaussian vector of zero mean and unit variance is generated. The vector of observations spatially correlated is obtained by multiplying y by the matrix of correlation.[2] Now we range the distances in increasing order in the vector h, and we create the vector of theoretical correlations associated to each component of h. Now, for each distance h_i, we can compute an observed correlation using a pair of associated observations. By observed correlation, we mean a realization of the variable $Z(h) = Y(s)Y(s+h)$. For the estimation, we use Bartlett kernel

$$k(x) = \begin{cases} 1 - |x|, & \text{if } |x| \leq 1 \\ 0 & \text{if not} \end{cases}$$

[2]Here, we use eigen value decomposition.

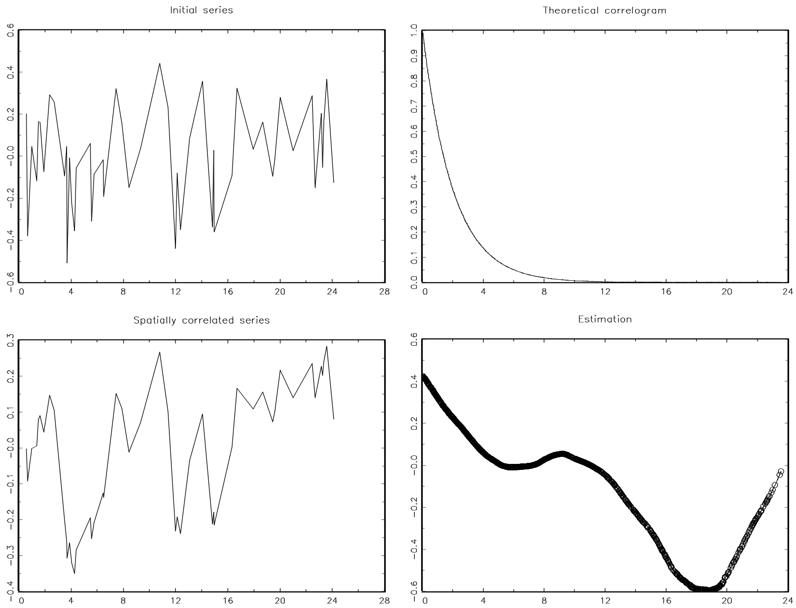

Figure 1: Graphs of simulations

Estimation results are plotted in Figure 1. The figure shows graphs of initial series, the theoretical correlogram, the spatially correlated series and the estimation. Simulations are conducted for 500 replications from our fixed location DGP for sample lengths of 30 observations. Overall, compared to the theoretical correlogram, our estimator seem to perform well in these experiments. However, edge location observations may seems to matter.

Appendix proofs

Proof of Proposition 1

By Assumption 2 we have $\Delta(n) \to 0$. Also, define

$$\lim_{n\to\infty} \Psi_n^\tau = \{(i,i') \in I \times I / \, |i-i'| = \tau, \tau \in \Lambda\} \equiv \Psi_\infty^\tau, \tag{11}$$

where we recall that $\Lambda = \{\tau = |i-i'|/(i,i') \in I\}$. Since $i - i' = \tau + \delta_i \in B(\tau, \Delta(n))$, where δ_i is a small variation (an increment), and $B(.)$ is an open ball. It follows that $\lim_{n\to\infty} |\delta_i| \leq \lim_{n\to\infty} \Delta(n) = 0$. That is, as given in Assumption 3, by definition of function $K(.)$, $\lim_{n\to\infty} K(\delta_i) = 1$. It then follows that

$$\lim_{n\to\infty} \hat{C}(\tau) = \sum_{(i,i')\in\Psi_\infty^\tau} U(i)U(i'). \tag{12}$$

To complete the proof, it remains to be shown that one can select $\Delta(n)$ such that the expectation of the cardinality of Ψ^τ_∞ converges towards infinity for $n \to \infty$. The conclusion will follow from the law of large numbers. Let $\Lambda^\tau_n = B(\tau, \Delta(n)) \cap \Lambda$, that is, the set of sample distances between τ and $\tau \pm \Delta(n)$. Since points coordinates are uniformly distributed, so are the associated distances. As a result, the expectation of $\#\Lambda^\tau_n$ is given by

$$E\left(\#\Lambda^\tau_n\right) = \#\Lambda \frac{\mu\left(B(\tau, \Delta(n))\right)}{\mu\left([a, b]\right)}, \tag{13}$$

where $\mu(.)$ denotes Lebesgue's measure and $\#$ is the cardinality symbol. The cardinality of $\#\Lambda$ is computed as

$$\#\Lambda = \binom{n}{2} = \frac{n!}{2!(n-2)!} = \frac{n(n-1)}{2}. \tag{14}$$

Now observe that $\mu(B(\tau, \Delta(n)))$ is the size of $B(\tau, \Delta(n))$, that is $2\Delta(n)$, and $\mu([a, b])$ denotes the length of $[a, b]$, that is $b - a$. Then, we obtain

$$E\left(\#\Lambda^\tau_n\right) = \frac{n(n-1)}{2} \frac{2\Delta(n)}{b-a}. \tag{15}$$

Finally, one can choose for example $\Delta(n) = 1/\sqrt{n}$ which implies that

$$\#\Psi^\tau_\infty = \lim_{n\to\infty} E\left(\#\Lambda^\tau_n\right) = +\infty. \square \tag{16}$$

Proof of Proposition 2

For notational convenience, we will suppress the set under the sum symbol. For $i = i'$ we have $\tau = 0$. Then $C_n(0) = \sum K(0)U(i)U(i')$. By Assumption 3, $K(0) = 1$. Then $C_n(0) = \sum_i U^2(i)$. One deduces that $V_n = \sum_i U^2 + \sum_{(i,j)} \omega_{ij}(.)(\sum_{(i,i')} K(.)U(i)U(i'))$. To show that $\hat{V}_n$ converges in probability towards V_n, let us apply Chebychev's inequality, that is for every $\eta > 0$,

$$P\left[\|\hat{V}_n - V_n\| > \eta\right] \leq \frac{E\|\hat{V}_n - V_n\|^2}{\eta^2}. \tag{17}$$

The job turns out to prove that $E\|\hat{V}_n - V_n\|^2 = o_p(1)$. To show this, decompose $\hat{V}_n$ and V_n as $\hat{V}_{n,1} = \hat{C}_n(0)$, $\hat{V}_{n,2} = \hat{C}_n(\tau)$, $V_{n,1} = C_n(0)$ and $V_{n,2} = C_n(\tau)$. Then we have

$$\begin{aligned} \left\|\hat{V}_n - V_n\right\| &= \left\|\hat{V}_{n,1} + \hat{V}_{n,2} - (V_{n,1} + V_{n,2})\right\|, \\ &= \left\|(\hat{V}_{n,1} - V_{n,1}) + (\hat{V}_{n,2} - V_{n,2})\right\|. \end{aligned} \tag{18}$$

Using the triangular inequality we obtain

$$\left\|\hat{V}_n - V_n\right\| \leq \left\|\hat{V}_{n,1} - V_{n,1}\right\| + \left\|\hat{V}_{n,2} - V_{n,2}\right\|. \tag{19}$$

Note that $\hat{V}_{n,1} \xrightarrow{\text{p}} V_{n,1}$ since $C_n(0)$ is a constant limit for $\hat{C}_n(0)$. From Proposition 1, we know that $\hat{V}_{n,2} \xrightarrow{\text{p}} V_{n,2}$. From the last part of (18) and using Minkoswski inequality we have

$$\begin{aligned} \left(\sum_{i=1}^{n} \left\|\hat{V}_{n,i} - V_{n,i}\right\|^2/n\right)^{1/2} &= \left(\sum_{i=1}^{n} \left\|(\hat{V}_{n,1i} - V_{n,i1}) + (\hat{V}_{n,2i} - V_{n,i2})\right\|^2/n\right)^{1/2} \\ &\le \left(\sum_{i=1}^{n} \left\|\hat{V}_{n,1i} - V_{n,1i}\right\|^2/n\right)^{1/2} \\ &+ \left(\sum_{i=1}^{n} \left\|\hat{V}_{n,2i} - V_{n,2i}\right\|^2/n\right)^{1/2} \\ &= o_p(1) + o_p(1) = o_p(1). \end{aligned}$$

The conclusion follows as we have shown that each term of the inequality converges in norm L_2 as $n \to \infty$. □

References

Anselin, L., and A. Bera (1998): "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics," *in Handbook of Applied Economic Statistics, A. Truchmuche (eds.)*, pp. 237–289.

Chen, X., and T. Conley (2001): "A New Semiparametric Spatial Model for Panel Time Series," *Journal of Econometrics*, (105), 59–83.

Conley, T. (1999): "Generalized Method of Moments Estimation with Cross Sectional Dependence," *Journal of Econometrics*, (92), 1–45.

Cressie, N. (1991): *Statistics for Spatial Data.* New York: Wiley-Interscience.

Domowitz, I., and H. White (1984): "Nonlinear Regression with Dependent Observations," *Econometrica*, (52–1), 143–161.

Driscoll, J. C., and A. C. Kraay (1998): "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data," *The Review of Economics and Statistics*, (80–4), 549–560.

Hansen, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, (50), 1029–1054.

Härdle, W. (1990): *Applied Nonparametric Regression.* Cambridge University Press.

Kelejian, H., and I. Prucha (1998): "A Generalized Spatial two-stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbance," *Journal of Real Estate Finance and Economics*, (17–1), 99–121.

——— (1999): “A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model,” *International Economic Review*, (40), 509–533.

Sajjan, S. G. (2000): “A note on central limit theorems for lattice models,” *Journal of Statistical Planning and Inference*, (83), 283–290.